International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 12 August 2023 **Revised:** 29 September 2023 **Accepted:** 16 October 2023*

Ensemble-Based Machine Learning Approach for Real-Time Person Counting in an Instant Attendance System

[1]Kumar Janardan Patra, [2,✉]Satyaprakash Swain, [3,✉]Suvendra Kumar Jayasingh, [4]Kodanda Dhar Naik, [5]Soumya Ranjan Prusty, [6]Sonalee Panda

[1]Department of Computer Science and Engineering, Institute of Management and Information Technology, Cuttack, BPUT, Odisha, India
[1]janardanpatra1997@gmail.com
[2]Department of Computer Science and Engineering, Institute of Management and Information Technology, Cuttack, BPUT, Odisha, India
[2]satyaimit@gmail.com
[3]Department of Computer Science and Engineering, Institute of Management and Information Technology, Cuttack, BPUT, Odisha, India
[3]sjayasingh@gmail.com
[4]Department of Computer Science & Engineering, Gandhi Institute of Engineering and Technology University, Gunupur, Odisha, India
kodandadhar.naik@giet.edu
[5]Vihave.ai
[5]soumya777prusty@gmail.com
[6]Vihave.ai
[6]sonalee1001@gmail.com
✉Corresponding Author

Abstract—Real-time attendance systems have become indispensable in various domains, including educational institutions and workplaces, as they automate attendance tracking and improve efficiency. This paper introduces a robust real-time attendance system that combines OpenCV and the You Only Look Once (YOLO) model. By integrating computer vision and deep learning techniques, the system achieves accurate and rapid face detection and recognition. Our proposed system utilizes OpenCV, a powerful computer vision library, to capture video streams from cameras. The YOLO model, a cutting-edge real-time object detection algorithm, is employed to identify and localize faces within the video frames. Thanks to YOLO's efficiency, the system ensures real-time processing, enabling seamless attendance recording. To enhance accuracy, the system employs a two-step approach consisting of face detection and face recognition. During the face detection phase, the YOLO model detects bounding boxes around faces. Subsequently, the system matches these detected faces against a pre-existing database of enrolled individuals using face recognition techniques. To improve performance, transfer learning techniques are applied to fine-tune the YOLO model on a diverse dataset containingvarious face images. This adaptation process ensures high precision and recall rates, even in challenging conditions such as varying lighting and occlusion. Experimental results demonstrate the effectiveness of the proposed real-time attendance system, achieving a high accuracy rate suitable for practical applications. Its real-time performance allows for seamless integration into existing attendance management workflows, resulting in time savings and improved administrative processes. The core intention of this paperwork is to develop a GUI page using Python programming, which will display the number of students who are present and number of students who are absent along with the number of students when section details are entered manually. In addition to this, pictures of the students who are present are displayed on the framework. It is called the real-time face detection which is very beneficial for managingacademic activity.

Keywords-Real-time Attendance System, OpenCV, YOLO model, Face Detection, Face Recognition, Transfer Learning, Computer Vision, Deep Learning, Attendance Management.

I. Introduction

In today's rapidly evolving world, the need for efficient attendance management systems has become increasingly crucial across various domains, including educational institutions and corporate workplaces. Traditional manual methods of attendance tracking often prove to be time consuming, error-prone, and inadequate for handling real time data. To overcome these challenges, the integration of cutting-edge technologies such as computer vision and deep learning has opened doors to advanced real-time attendance systems. This research paper introduces a robust real-time attendance system that harnesses the capabilities of OpenCV, a powerful computer vision library, and the You Only Look Once (YOLO) model, an exceptional real-time object detection algorithm. By seamlessly combining these technologies, our proposed system aims to revolutionize the way attendance tracking is conducted. The foundation of our real-time attendance system lies in OpenCV, which acts as the primary tool for capturing live video streams from cameras. Leveraging its comprehensive set of functions, the system efficiently processes each frame and extracts vital information from the video feed. To achieve real-time face identification and localization, we employ the YOLO model, renowned for its high-speed object detection capabilities. Through the adaptation of YOLO to our specific attendance context and fine-tuning it with transfer learning, our system ensures accurate and prompt face detection, even in dynamic environments. The system operates through a two-step process: first, it employs the YOLO model to detect bounding

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *12 August 2023* ***Revised:*** *29 September 2023* ***Accepted:*** *16 October 2023*

---

boxes around faces, and subsequently, it utilizes face recognition techniques to match these detected faces against a pre-existing database of enrolled individuals. This comprehensive approach guarantees both real-time performance and high accuracy, eliminating the risk of false identifications. Extensive experimental validation demonstrates the effectiveness of the proposed real-time attendance system. Its remarkable precision and recall rates, coupled with seamless integration into existing attendance management workflows, position it as an ideal solution for a wide range of practical applications. This research paper presents the development, implementation, and evaluation of the real-time attendance system, showcasing its potential to revolutionize attendance tracking and management in educational institutions, workplaces, and beyond. By harnessing the combinedpower of OpenCV and the YOLO model, we introduce aninnovative approach to address attendance challenges andenhance administrative efficiency.

II. YOLO (YOU ONLY LOOK ONCE)

YOLO (You Only Look Once) is a widely used object detection algorithm known for its real-time performance and accuracy. It treats object detection as a regression problem and utilizes convolutional neural networks (CNNs) to predict object classes and bounding box coordinates in a single pass. YOLO divides the input image into a grid and processes each grid cell to detect objects, taking into account global context information. Its unified approach, speed, and strong feature representation contribute to its effectiveness. YOLO is flexible and can adapt to different detection scenarios, making it suitable for a variety of applications. Its fast and accurate object detection capabilities have made it a popular choice in the field.

A. Architecture of YOLO

The YOLO (You Only Look Once) architecture shares some similarities with Google Net. It consists of a total of 24 convolutional layers, four max-pooling layers, and two fully connected layers. These layers are arranged in a sequential manner to process the input image and extract relevant features as shown in fig 1. However, it's important to note that the specific configuration and design of YOLO may vary depending on the version and variant being referred to.

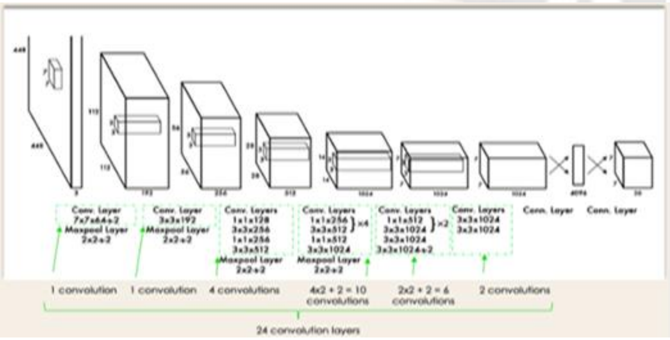

Figure. 1. YOLO Architecture

The YOLO architecture operates as follows:

- The architecture starts with a 1x1 convolutional layer, which is used to reduce the number of channels in the input.
- This is followed by a 3x3 convolutional layer that generates a cuboidal output.
- Throughout the architecture, the Rectified Linear Unit (ReLU) activation function is applied to introduce non-linearity, except for the final layer, which uses a linear activation function.
- To enhance the model's performance and prevent overfitting, additional techniques like batch normalization and dropout are employed. Batch normalization normalizes the outputs of each layer, improving training stability and accelerating convergence. Dropout randomly drops out a fraction of the neurons during training, reducing the model's reliance on specific features and enhancing its generalization ability.

III. EXISTING WORKS

Nagoriya et al. [1] emphasizes the importance of finding an innovative solution to streamline attendance management, aiming to reduce time, effort, and resources involved in the process. The proposed system utilizes face recognition technology to detect and identify the faces of students or employees by extracting distinct facial features from captured images. The system is implemented using a Raspberry Pi, enabling fast and accurate detection and recognition of human faces from images or videos captured by a camera. The implementation leverages the OpenCV image processing library and ensuring independence from specific hardware or software requirements. Okechukwu M. Chukwude et al. [2] conducted research focusing the development and testing of a web application named Roll Call for attendance management in the Faculty of Engineering at the University of Ilorin. The system leverages face recognition technology implemented using Python, OpenCV, and Sci-kit Learn to accurately and efficiently mark attendance. Students can enroll in courses, upload their face data, and access attendance records for their enrolled courses. The web interface of Roll Call is developed using HTML5, the Twitter Bootstrap CSS framework, and JavaScript, ensuring a user-friendly experience for both lecturers and students. Khaled Elabbani et al. [3] conducted research and proposes an automated attendance system for exams using face recognition technology. The system employs the LBP algorithm for facial recognition and utilizes the HaarCascade and MTCNN algorithms for face detection. The system exhibits high accuracy in identifying students, even in challenging conditions such as varying lighting and changes in facial appearance. Kristina Ranči ć et al. [4] conducted research with title "Animal Detection and Counting from UAV Images Using Convolutional Neural Networks" which presents a study on utilizing small unmanned aerial vehicles (UAVs) and deep learning techniques for the detection and counting of deer in north-western Serbia. The authors compare the performance of various network architectures, including versions of You Only Look Once (YOLO) and Single Shot Multibox Detector

(SSD), in detecting deer in dense forest environments. The models are trained on a manually annotated dataset and evaluated using metrics such as mean average precision (mAP), precision, recall, and F1 score. Ashwin Rao et al. [5] introduces a novel approach to streamline attendance marking in educational institutions through real-time face recognition. Here deep learning (DL) models are used to detect face of student. It eliminates the need for manual intervention by automatically analyzing and tracking attendance. Chowdhury et al. [6] have presented a novel automatic student attendance system that employs Convolutional Neural Networks (CNN) and face recognition. The system utilizes face detection and recognition techniques to accurately identify multiple individuals from a video stream. Rasool et al. [7] have introduced system utilizes facial recognition powered by Artificial Intelligence to match captured photos with those stored in its database, enabling automatic attendance marking. The system incorporates CNN [4] and the pre-trained FaceNet model, along with the Multi-Task Cascaded Convolution Neural Network (MTCNN) for accurate face detection. Sailendra K et al. [8] have proposed a facial recognition system that utilizes a Convolution Neural Network (CNN) model to achieve fast and accurate face identification in various domains. They emphasize the practical applications of face recognition in domains such as information security, biometrics, access control, law enforcement, smart cards, and surveillance systems. It demonstrates the success of their approach by leveraging CNNs, particularly in real-time systems where complete images serve as input. They outline the key steps involved, such as feature selection, feature extraction, and training [8]. Kliangsuwan, T. et al. [9] conducted experiments in three different scenarios to evaluate the proposed system: online classes, on-site classes, and a standard dataset with complex cases. The obtained results demonstrate excellent accuracy across all scenarios, with the online scenario achieving a remarkable recognition accuracy of 100%. The on-site scenario achieved an accuracy of 98.29%, while the standard-dataset scenario achieved 92.50% accuracy, leveraging deep learning (DL) techniques. it can be utilized on portable devices such as tablets and smart phones Edison Kagona et al. [10] focuses on the implementation of a facial recognition attendance system using OpenCV and deep learning techniques. It discusses the use of face embeddings for deep metric learning, including the steps of face detection, feature extraction, and face comparison. The authors also describe the development process of the proposed system, including studying the existing system, gathering requirements, and implementing the interfaces using HTML, Bootstrap, and Django. The Fast-Yolo-Rec algorithm is proposed by ZAREI.Net al. [11]. In this study a solution for real-time vehicle detection with enhanced speed and accuracy. They introduce a Yolo-based detection network and LSTM-based position prediction networks. The algorithm incorporates a spatial semantic attention mechanism (SSAM) to improve accuracy and speed. By employing alternating cycles of detection and prediction, the algorithm achieves faster vehicle position detection without compromising accuracy Mane. V et al. [12] introduce a system employs real-time face detection and recognition using the OpenCV library. By utilizing a dedicated camera module, the system captures real-time images of a classroom and evaluates student attendance based on their data. Attendance information is subsequently sent via email to the relevant individual. The paper presents a novel solution that combines face recognition technology, real-time image processing and IoT integration to enhance attendance management in an accurate and secure manner. Suhane. A K et al. [13] examines the utilization of the YOLO model for human detection and crowd counting tasks in computer vision. It highlights the practical applications of these tasks in various domains, including surveillance, security, crowd management, and traffic analysis. The paper acknowledges the remarkable achievements of deep learning models, specifically YOLO, in delivering efficient and accurate object detection and counting capabilities. While praising YOLO's speed, efficiency, and state-of-the-art performance, the paper also addresses its limitations. Parida et al. [17] have discussed a novel approach for recognition of human activity using vision based method.

IV. RESEARCH DESIGN

A. *Why Does YOLO Outperform Object Detection Algorithm*

YOLO (You Only Look Once) is a popular object detection algorithm known for its real-time performance and high accuracy. It outperforms other object detection algorithms due to several reasons. These are Unified approach, Speed, Contextual information, Strong feature representation, Training strategy, Flexibility. While YOLO has many advantages, it's important to note that other object detection algorithms like Faster R-CNN and SSD (Single Shot Multi Box Detector) excel in certain aspects or domains. The choice of the best algorithm depends on the specific requirements of the task at hand.

- Classification-based algorithms
- Regression-based algorithm

B. Versions of YOLO Models

It is important to note that developments in the field of computer vision and deep learning are ongoing, and new versions or variations of the YOLO model may have been released since then. It's always a good idea to consult the latest research and resources to ensure you have the most up-to-date information on the YOLO model or any other deep learning models.

C. *YOLO V3*

YOLO V3 is the most popular, fast, and accurate real-time object detection algorithm, including multi-scale prediction and darknet53 as backbone classifier. It also uses a convolution neural network for object identification and detection, which is also called darknet53. It can detect multiple objects in one frame and is efficient in finding the location of objects in an image. In this process, the image is divided into n x n. Grid cell, and for each cell and set of cells, it predicts multiple objects in the image by considering the cell probability, predicts bounding boxes over images, and chooses the best one with more probability of having an object in an image. YOLO V3 can accept different shapes of images where the shape should be divided by 32 without any remainder, such as 416 x 416, 608 x 608, and so on.

YOLO V3 accepts the input in a batch of images where the input shape is of the form (n x w x h x c) where • n represents no of images

- w represents the width of the image
- h represents the height of the image
- c represents the channel of the image

YOLO V3 makes use of 1 x 1 kernel on the down-sampled image at 3 different detection layers, and it produces an output of shape (n x n x c) Where (n x n) represents the feature map size after the formula calculates down-sampled and c

$$(b * (5 + c))$$

Where b is a number of detection layers and c is no classes.

The proposed system generates (3 * (5 + 13)) as it tries to predict 13 classes and 3 different layers.

D. *YOLO V5*

YOLO V5 is the most advanced and latest object detection algorithm based on modified Convolutions Neural Networks (CNN) architecture that uses Cross Stage Partial CSPDarknet as the backbone with Spatial Pyramid Pooling (SPP) layer.The YOLO V5 architecture is divided into 3 parts; in the first part, CSP-Darknet is used to extract the valuable characteristics of input images. The second part of the architecture, called Neck and YOLO V5, uses PANet, a pyramid approach that helps to identify the same Object at different scales. The third part of architecture is called the Head, which is responsible for final object identification from an input image and draws anchor boxes on images to construct the final output with class labels and class names. YOLO V5 builds upon the success of previous versions and adds several new features and improvements. Unlike YOLO, YOLO V5 uses a more complex architecture called EfficientDet (architecture shown below), based on the EfficientNet network architecture as shown in fig.2. Using a more complex architecture in YOLO V5 allows it to achieve higher accuracy and better generalization to a wider range of object categories.

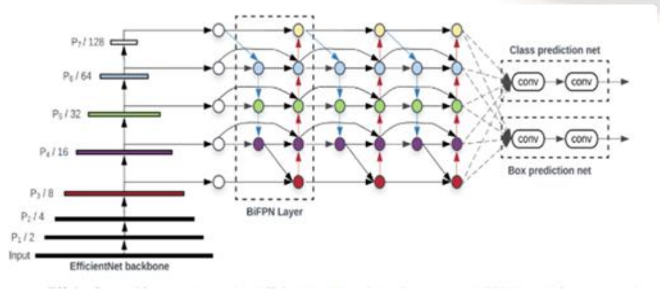

Figure. 2. Architecture of the EfficientDet model

E. *YOLO V8*

The most recent model in the YOLO family is called YOLO V8. You Only Look Once is the abbreviation for this family of models, which are so named because they can accurately forecast every object in an image with a single forward pass. YOLO models are pre-trained on huge datasets such as COCO and ImageNet. This gives them the simultaneous ability to be the Master and the Student. They provide highly accurate predictions on classes they are pre-trained on (master ability) and can also learn new classes comparatively easily (student ability). YOLO models are also faster to train and have the ability to produce high accuracy with smaller model sizes. They can be trained on single GPUs, making them more accessible to developers. YOLO V8 is the latest iteration of these YOLO models (as of early 2023). It has undergone a few major changes from its ancestors, such as anchor-free detection, the introduction of C3 convolutions, and mosaic augmentation.

F. *YOLO NAS*

YOLO NAS refers to the integration of Neural Architecture Search (NAS) techniques with the object detection algorithm. NAS is a method that automates the design of neural network architectures by searching for the optimal network architecture for a given task. By leveraging NAS, YOLO NAS aims to improve the efficiency and effectiveness of the YOLO algorithm by automating the process of architecture design.

V. METHODOLOGY.

A. Proposed System Flow

We employ three versions of the YOLO (You Only Look Once) model, namely YOLO V5, YOLO V8, and YOLO V8 NAS, to conduct our research. These models have been selected for their significant advancements in object detection and person counting capabilities. The YOLO V5 model is renowned for its efficiency and high accuracy in object detection tasks. It incorporates advanced architecture and optimization techniques to achieve real-time performance without compromising detection quality. By utilizing the YOLO V5 model, we aim to leverage its improved speed and precision to enhance person counting in attendance systems, specifically in meeting scenarios. In addition to YOLO V5, we integrate the YOLO V8 model into our research methodology. Building upon the successes of earlier YOLO versions, the YOLO V8 model further improves detection accuracy through advanced network architectures and optimization techniques. By incorporating the YOLO V8 model, we seek to explore its potential for achieving more accurate and robust person counting in attendance systems.Furthermore, we incorporate the YOLO V8 NAS model, which utilizes Neural Architecture Search. This approach automates the process of searching for optimal neural network architectures for object detection tasks. By employing the YOLO V8 NAS model, we aim to further enhance the accuracy and efficiency of person counting in attendance systems, taking advantage of its automatic architecture search capability. By utilizing these three versions of the YOLO model (YOLO V5, YOLO V8, and YOLO V8 NAS), our research aims to thoroughly investigate their strengths and capabilities in achieving accurate person counting in meeting scenarios within attendance systems. This approach allows us to compare and contrast their performances, ultimately identifying the most suitable model or combination of models to enhance the accuracy and robustness of attendance monitoring.

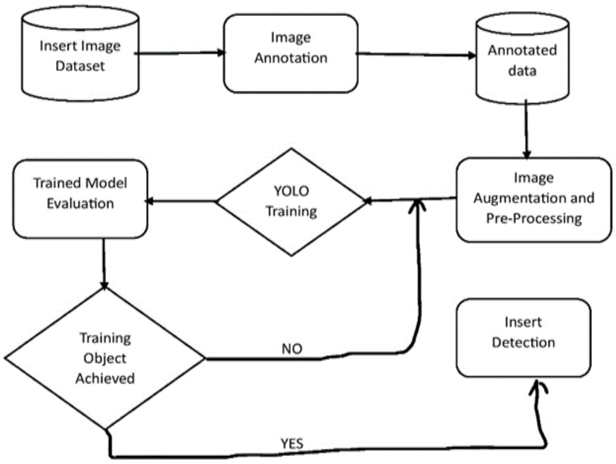

Figure. 3. Data Preparation for Model Training

In our proposed system, we follow a systematic flow to develop an accurate person counting model using YOLO (You Only Look Once) variants. First, we collect a dataset of images or videos capturing meeting scenarios where person counting is required for attendance monitoring. The data undergoes preprocessing steps [13], including resizing, normalization, and augmentation, to ensure consistency and enhance model performance. Next, we select the appropriate YOLO model variant, such as YOLO V5, YOLO V8, or YOLO V8 NAS, based on their strengths in object detection and person counting tasks, as shown in above fig. 3.

B. Data Gathering

Dataset is created by collecting a total of 300 images from the classrooms of the Institute of Management and Information Technology, located in Cuttack. The dataset consisted of images capturing both students and faculty members within the classroom environment. The primary objective was to develop a robust model capable of accurately detecting and classifying individuals in an educational setting which is shown in fig.4.

Figure.4 Sample of Image Dataset

The collection process involved capturing images from various angles and perspectives to account for different poses and variations in lighting conditions. Care was taken to ensure the representation of a diverse range of individuals, including students of different ages, genders, and ethnicities, as well as faculty members from various academic disciplines.

C. Data Preparation

The dataset consisting of 300 images was collected from Institute of Management and Information Technology, Cuttack, encompassing both students and faculty members. To ensure the dataset's efficacy, a meticulous annotation and labeling process was undertaken using specialized software. During annotation, skilled annotators manually drew bounding boxes around individuals in the images with precision.

D. How Does YOLO Object Detection Work?

YOLO algorithm aims to predict a class of an object and the bounding box that defines the object location on the input image. It recognizes each bounding box using four numbers:

- Centre of the bounding box (bx,by)
- Width of the box (bw)
- Height of the box (bh)

In addition to that, YOLO predicts the corresponding c number for the predicted class as well as the probability of the prediction Pc.

1) Bounding Boxes

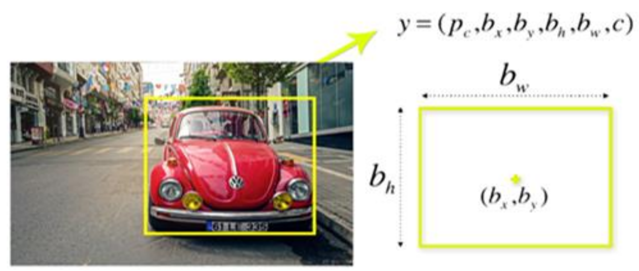

Figure. 5. Bounding Boxes parameters

In this example, we have an image containing an object called "vehicle" that we want to detect. To create a bounding box around this object, which is shown in fig 5, we need four parameters: bx (bounding box x-coordinate), by (bounding box y-coordinate), bw (bounding box width), and bh (bounding box height). The bx and by values indicate the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 12 August 2023 **Revised:** 29 September 2023 **Accepted:** 16 October 2023*

centre of the bounding box. By identifying the centre, we can draw a bounding box around the object. The width of the bounding box is represented by bw, and the height is represented by bh. Using these four parameters, we can draw the bounding box.

2) Grid Cell Concept:

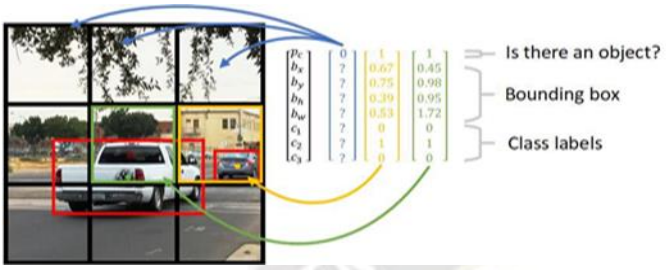

Figure. 6. Bounding Boxes parameters with Grid Cell

As per its concept YOLO creates a 19x19 grid on the image, but here we create a 3x3 grid cell for understanding, as seen in fig 6. As a result, we have divided the image into 9 sections. The first element in Fig 6 will be 1 if an object is present and 0 if not, followed by the values bx, by, bh, and bw, and the last three will represent the class present in that cell based on the classes we have. We have a piece of the automobile just like in the middle cell; hence the second value is 1.

3) Anchor Boxes

To detect and localize various objects in an image, state-of-the-art object detection models, like YOLO, utilize anchor boxes as a prior. The collection of bounding boxes with predefined height and width is referred to as "anchor boxes." These anchor boxes are used to filter predictions based on class scores, perform non-maximum suppression to eliminate redundant detections, and prioritize accurate localization through intersection over union (IOU) calculations.

Figure. 7. Anchor Boxes

The predefined anchor boxes are tiled across the image based on the number of classes of bounding boxes. For each anchor box, the goal is to determine which object bounding box has a higher intersection-over-union (IOU) value. In Fig 7, the first three anchor boxes are empty, indicating that no object is detected with a high IOU. However, the fourth anchor box contains an object resembling an airplane, and ithas a higher IOU with the corresponding object bounding box. The objective is to assign anchor boxes based on the IOU value.

To facilitate model development and evaluation, the dataset was divided into three subsets: training, testing, and validation. Approximately 70% of the data was allocated for training to allow the model to learn from a substantial portion. The testing set, comprising 20% of the data, served as an independent evaluation set to assess the model's generalization capabilities. The remaining 10% was allocated to the validation set, aiding in the fine-tuning of model parameters and mitigating the risk of over fitting.

VI. EVALUATION METRICS

Comparing between models we used some evaluation metrices that are MAP (mean Average Precision), precision, and recall used to assess the performance of object detection or recognition models.

A. mAP

"mAP" stands for Mean Average Precision. It is a commonly used evaluation metric to assess the performance of object detection models, including YOLO. The YOLO model detects objects in an image by dividing the image into a grid and predicting bounding boxes and class probabilities for each grid cell. During evaluation, the predicted bounding boxes are compared to the ground truth bounding boxes to calculate the precision and recall values. The precision-recall curve is then created by varying the confidence threshold for accepting detection.

mAP is calculated by averaging the average precision (AP) values obtained at different recall levels. Higher MAP values indicate better performance, with a maximum value of 1. mAP is often used to compare different versions of YOLO models or to compare YOLO with other object detection models to determine which one performs better on a given dataset.

1. Calculate Average Precision (AP) for each class
2. Calculate Mean Average Precision (mAP)
3. Calculate Average Precision (AP) for each class
   Let TP(R) be the number of true positives at a given recall level R, and FP(R) be the number of false positives at the same recall level.
   Precision P(R) at recall level R is calculated as:

$$P(R) = TP(R) / (TP(R) + FP(R))$$

   Recall R is typically varied from 0 to 1 and precision values are computed at different recall levels. Compute precision and recall values for each class at different recall levels, then calculate the area under the precision-recall curve (AP) for each class.

4. Calculate Mean Average Precision (mAP):

Average the AP values obtained for each class to obtain the Mean Average Precision (mAP).
Let N be the total number of classes.

$$MAP = (AP\_1 + AP\_2 + ... + AP\_N) / N$$

Where AP_1, AP_2, ..., AP_N are the Average Precision values calculated for each class.

B. *Precision*

Precision is calculated by dividing the number of true positives (TP) by the sum of true positives and false positives (FP):

Precision = TP / (TP + FP)

True positives (TP) represent the correctly detected objects, while false positives (FP) represent the objects that were incorrectly detected or falsely identified.

C. *Losess*

There are three primary components that contribute to the loss calculation: box loss, class loss, and object loss. These components help optimize the model's performance in object detection tasks.

- Box Loss: The box loss, also known as the localization loss, measures the discrepancy between the predicted bounding box coordinates and the ground truth bounding box coordinates. It quantifies how accurately the model localizes the objects of interest. The box loss is typically calculated using regression-based loss functions such as Smooth L1 loss or Mean Squared Error (MSE).
- Class Loss: The class loss evaluates the model's ability to assign correct class labels to detected objects. It measures the difference between the predicted class probabilities and the ground truth class labels. The class loss is commonly computed using cross-entropy loss or softmax loss functions.
- Object Loss: The object loss reflects the confidence or certainty of the model in detecting objects within a bounding box. It quantifies the discrepancy between the predicted objectness score (indicating the presence of an object) and the ground truth objectless score. The object loss is typically calculated using binary cross-entropy loss.

VII. EXPERIMENT

A. *Experiment of YOLO V5*

YOLO V5 Deep Sparse is a variant of the YOLO V5 object detection algorithm that has been specifically designed to achieve higher accuracy and real-time performance for identifying persons[14] from dataset. In this experimentation, the algorithm is trained with various parameters, including input image size, epochs, and batch size, to optimize its performance. In order to detect the Object using YOLO V5, we have to clone the pre-trained YOLO V5 model to use the weights for training the model on the custom dataset. To evaluate the model's performance, the mAP metric is employed. mAP measures the precision-recall trade-off and provides an overall assessment of the model's accuracy in detecting and localizing objects. At epoch time the below given pattern are created which are shown in fig. 8.

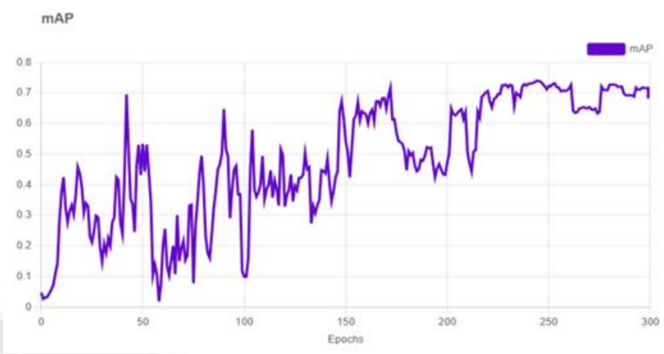

Figure. 8. mAP pattern created at time of Epoch

During training, the contributions of these three losses are combined to compute the overall loss which is shown in fig. 9. The relative weights assigned to each loss component may vary depending on the specific implementation and task requirements. By optimizing the overall loss through back propagation and gradient descent, the model learns to improve its object detection performance, including accurate localization, correct class prediction, and reliable object presence estimation.

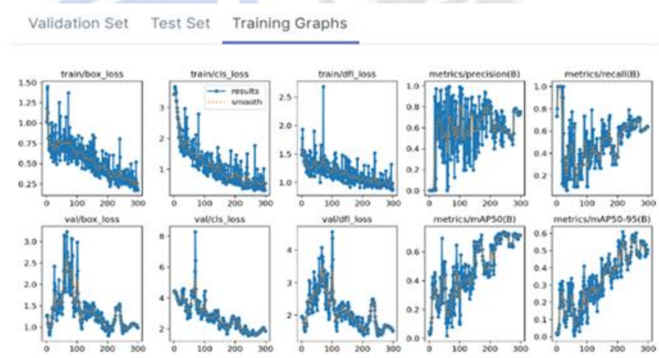

Figure. 9. Losses at time of training dataset of YOLO V5

In order to improve accuracy and real-time performance with higher FPS (Frame Per Second), the YOLO V5 Deep Sparse Baseline algorithm is utilized. This algorithm utilizes two methods - pruning and quantization. Pruning is a process of removing the unus ed weights in the model, thereby reducing its size and computational requirements.

TABLEI.YOLO V5 Experimental Result

Algorithm	Input	Epoch	Batch Size	Optimizer	mAP(.5)	mAP (.5-.95)
YOLO V5 ultra Latics	300	5	64	sgd	72.32 %	57.3 %
YOLOv5ultra Latics	300	10	32	sgd	79.23 %	64.45 %
YOLOv5Deep Sparse	300	5	32	sgd	73.34 %	58.49 %
YOLOv5Deep Sparse	300	10	64	sgd	79.5%	65.01 %

B. Experiment of YOLO V8

With an overview of the architecture of YOLO V4 which represents the latest advancements in the series. YOLO V8 is an enhanced version of the YOLO (You Only Look Once) object detection algorithm, specifically designed to improve accuracy and performance in various computer vision applications. In our experimentation, we employ YOLO V8 to address the task of object detection and person counting in meeting scenarios. The algorithm is trained using different parameters, such as input image size, number of epochs, and batch size, to optimize its performance on the given dataset. By experimenting with these parameters, we aim to find the optimal configuration that achieves the best results in terms of accuracy and efficiency. To evaluate the model's performance, we utilize the mAP (mean Average Precision) metric, which provides a comprehensive assessment of the model's ability to accurately detect and localize objects in the given dataset. mAP evaluate pattern given in fig. 10

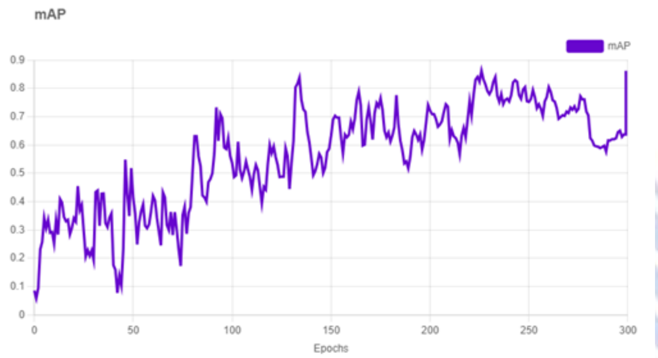

Fig. 10. Pattern at time of calculation for mAP

During training, the contributions of these three losses are combined to compute the overall loss which are shown in fig. 11. The relative weights assigned to each loss component may vary depending on the specific implementation and task requirements. By optimizing the overall loss through back propagation and gradient descent, the model learns to improve its object detection performance, including accurate localization, correct class prediction, and reliable object presence estimation.

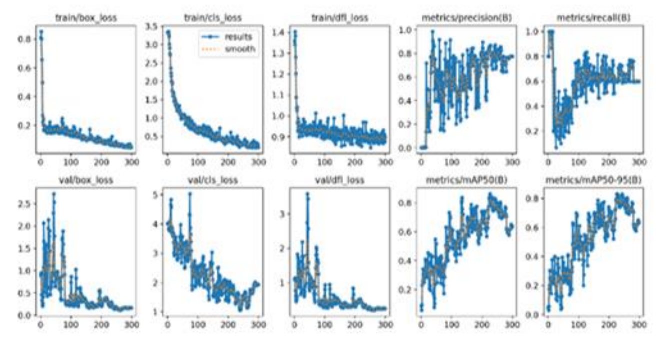

Fig. 11. Losses at time of training dataset of YOLO V8

After complete 10 epochs we found mAP is 86.3%. and also calculate precision i.e 85.9% and recall is 60.9% from our train model of YOLO V8. Those values are stored in table II.

TABLE II.YOLO v8 EXPERIMENTATION RESULT

Epoch	Batch	mAP(.5)	mAP(.5-.95)
10	8	81.55%	61.97%
20	8	86.32%	67.32%

C. Experiment of YOLO V8 NAS

YOLO V8 NAS (Neural Architecture Search) is an improved version of the YOLO family, and it comes with some additional features which provide a higher accurate result with a high frame rate. YOLO V8 NAS uses decoupled Head and leading label assignment strategies such as simOTA for large-scale SOTA results. For the building of YOLO NAS. YOLO V8 NAS uses darknet as a baseline backbone architecture for key feature extraction.

Object detection algorithms always perform two tasks: localization, which means locating the Object, and classification, classifying objects into particular classes. Neural Architecture Search (NAS), there is several approaches and techniques that have been developed to automate the search for optimal neural network architectures. These are: Reinforcement Learning (RL), Evolutionary Algorithms (EA), Gradient-based Optimization, Genetic Programming, Bayesian Optimization.

To evaluate the model's performance, we utilize the mAP (mean Average Precision) metric, which provides a comprehensive assessment of the model's ability to accurately detect and localize objects in the given dataset. mAP evaluate pattern given in fig.12, The mAP metric considers both precision and recall, providing a balanced measure of detection performance.

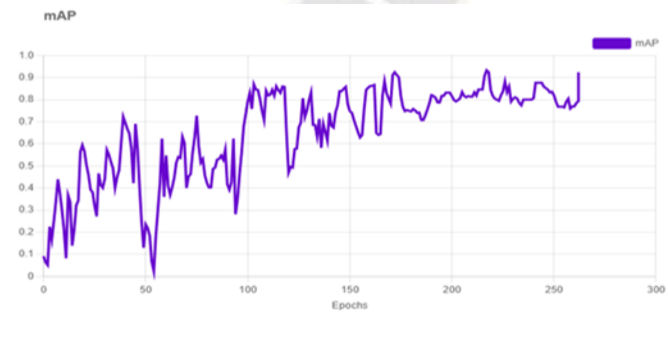

Figure. 12. Pattern at time of calculation of mAP in v8 NAS

In case of train YOLO V8 NAS model we also found some loss i.e. box loss, class loss, object loss when training our data set, pattern looks like, which is shown in below fig. 13. the YOLO V8 architecture introduces several modifications compared to previous YOLO versions.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 12 August 2023 **Revised:** 29 September 2023 **Accepted:** 16 October 2023*

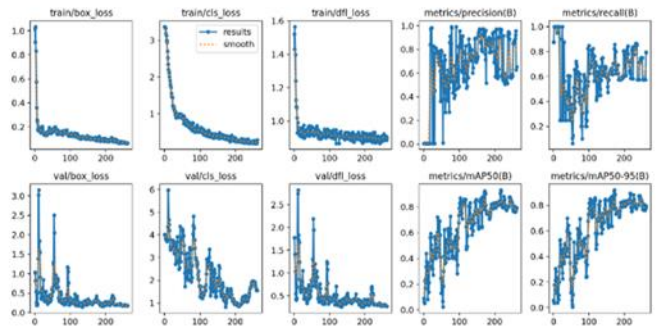

Figure. 13. Losses at time of training dataset of YOLO V8 NAS

TABLE III. YOLO V8 NAS EXPERIMENTATION RESULT

Epoch	Batch	Image size	mAP(.5)	mAP(.5-.95)
10	8	300	91.6%	85.8%
12	8	300	92.6%	89.3%

VIII. PERFORMANCE EVALUATION

From the result of Table IV, we can find that YOLO V8 NAS is performing very well compared to others model and graphical representation shown in fig. 14.

TABLE IV. EXPERIMENTATION RESULTS

Algorithm	Epoch	Batch	mAP(.5)	mAP(.5-.95)
YOLOv5	10	32	79.5%	65.01%
YOLOv8	10	32	86.32%	67.32%
YOLOv8 NAS	10	32	92.6%	89.3%

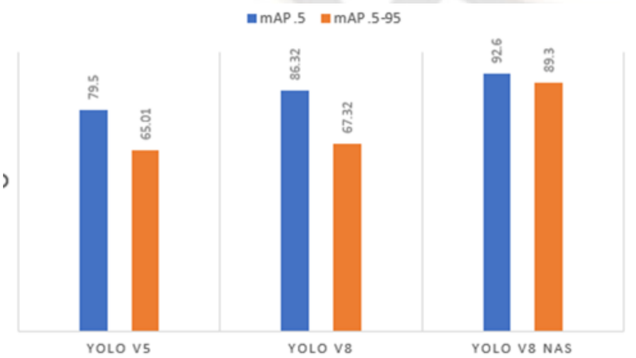

Figure. 14. Graphical representation of table IV

In Table-V it is shown that the mAP(mean absolute precision), Precision and Recall value are stored and between these taken three model are compare shown graphically in fig. 15. which is given below.

TABLE V. mAP, Precision and Recall values

Algorithm	Epoch	Batch	mAP	Precision	Recall
YOLO V5	10	32	79.5%	70.1%	73.9%
YOLO V8	10	32	86.32%	85.9%	60.9%
YOLO V8 NAS	10	32	92.6%	90.2%	82.5%

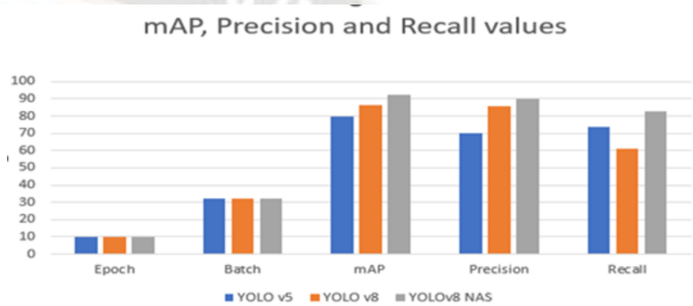

Figure. 15. Comparison table

IX. CONCLUSION

In the realm of attendance systems, researchers have approached the problem using various techniques, including classification methods[15-16] and computer vision-based object detection approaches. However, a significant portion of the existing literature lacks detailed explanations on how they arrived at their solutions and the modifications made to enhance real-time accuracy. Our literature survey revealed that many researchers did not explicitly outline their solution development process or the specific changes implemented to achieve their desired outcomes. This knowledge gap highlights the need for more comprehensive discussions and insights into the problem-solving methodologies employed in attendance system research. In our proposed solution, we have achieved promising results by utilizing the YOLOv8 NAS architecture. With a confidence threshold of 0.5, we obtained amAP of 92.4%, and by varying the confidence threshold from 0.5 to 0.95 with a step of 0.05, we achieved amAP of 89.3%. This demonstrates the effectiveness of our approach in achieving high accuracy in real-time attendance monitoring. Several factors contributed to the improved accuracy of our solution. Additionally, we addressed the issue of class imbalance by balancing the dataset, ensuring that each class had a sufficient number of samples for robust training. In the phase of data annotation, we have taken care of our data and its corresponding classes and using the ROBOFLOW annotation tool, we have annotated the dataset. We have utilized YOLO V8 and YOLO V8 NAS for proposed system experimentation in order to find efficient SOTA methods. In addition, as per our experimentation, we found that YOLO V8 NAS could perform better in terms of accuracy with 10 epochs. In contrast, YOLO V8 NAS performs much better than YOLO V8 and YOLO V5 among these. In future work, we aim to extend our proposed attendance system by incorporating real-time data from CCTV cameras. This expansion will allow us to capture attendance and count individuals in real-time scenarios, such as classrooms, meeting rooms, or other relevant environments. To accomplish this, we will develop a pipeline to receive video streams from CCTV cameras and process them in real-time using our attendance system.

ACKNOWLEDGEMENT:

This is a original academicals research work which is not funded by any financial agencies.

REFERENCES

[1] Nagoriya, H. (2020). Attendance System using Face Recognition utilizing OpenCV Image Processing Library. International Journal of Recent Advances in Science, Engineering and Technology (IJRASET), 8(6), ISSN: 2321-9653;

[2] Chukwude, O. M., Surajudeen-Bakinde, N. T., Zakariyya, S. O., Ogunsakin, J. B., Akanni, J., Olayanju, S. A., &Ehiagwina, F. O. (2023). Development of Face Recognition-Based Attendance System. Sule Lamido University Journal of Science & Technology, 6(1&2), 214-224. ISSN: 2736-0903

[3] Elabbani, K., Jafar, A., &Khalafullah, W. (2022). An attendance system for exams using face recognition Based on MTCNN and OpenCV algorithms. International Science and Technology Journal, 30, July.

[4] Ranˇci´c, K.; Blagojevi´c, B.; Bezdan, A.; Ivoševi´c, B.; Tubi´c, B.; Vraneševi´c, M.; Pejak, B.; Crnojevi´c, V.; Marko, O. Animal Detection and Counting from UAV Images Using Convolutional Neural Networks. Drones 2023, 7, 179.

[5] A. Rao, "AttenFace: A Real Time Attendance System Using Face Recognition," 2022 IEEE 6th Conference on Information and Communication Technology (CICT), Gwalior, India, 2022, pp. 1-5, doi: 10.1109/CICT56698.2022.9998001.

[6] S. Chowdhury, S. Nath, A. Dey, and A. Das, "Development of an automatic class attendance system using cnn-based face recognition," in 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE). IEEE, 2020, pp. 1–5

[7] Rasool, A., Roohi, S., & Majumder, A. (2020). Automated Attendance System using Few-Shot Learning Approach. In 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) (pp. 1-5). Presented at the Amity University, Noida, India. June 4- 5, 2020.

[8] S. Kumar, D. Kathpalia, D. Singh and M. Vats, "Convolutional Neural Network based Automated Attendance System by using Facial Recognition Domain," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 654-659, doi: 10.1109/ICICCS48265.2020.9121163.

[9] T. Kliangsuwan, A. Heednacram and K. Silanon, "Face Recognition Algorithms for Online and On-Site Classes," 2022 19th International Joint Conference on Computer Science and Software Engineering (JCSSE), Bangkok, Thailand, 2022, pp. 1-6, doi: 10.1109/JCSSE54890.2022.9836309.

[10] Edison Kagona (2022), Facial Recognition Attendance Scheme on CCTV Cameras Using Open Computer Vision and Deep Learning: A Case Study of International University of East Africa (IUEA). Advanced Journal of Science, Technology and Engineering 2(1), 1-27. DOI: 10.52589/AJSTEHYVTCZ.

[11] V. Mane, S. Indulkar, S. Khandade, O. Khelge, A. Mitkari. "Smart Attendance System using Image Processing and Raspberry PI." International Journal of Computer Research and Technology (IJCRT), vol. 9, no. 10, October 2021, pp. [page numbers]. ISSN: 2320- 2882.

[12] A.K. Suhane, A. Vani, H. Parihar, L. Saxena (2023). Human Detection and Crowd Counting using YOLO. Proceedings of Tech Fest IIT ROORKEE, IIT Roorkee.

[13] N. Zarei, P. Moallem and M. Shams, "Fast-Yolo-Rec: Incorporating Yolo-Base Detection and Recurrent-Base Prediction Networks for Fast Vehicle Detection in Consecutive Images," in IEEE Access, vol. 10, pp. 120592-120605, 2022, doi: 10.1109/ACCESS.2022.3221942.

[14] V. Mishra, S. Raj, T. Singhal and C. Sankhla, "Intelligent Face Recognition based attendance system," 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2022, pp. 1- 4, doi: 10.1109/ICRITO56286.2022.9964862

[15] Swain, S. ., Behera, N. ., Swain, A. K., Jayasingh, S. K. ., Patra, K. J. ., Pattanayak, B. K. ., Mohanty, M. N. ., Naik, K. D. ., & Rout, S. S. . (2023). Application of IoT Framework for Prediction of Heart Disease using Machine Learning. International Journal on Recent and Innovation Trends in Computing and Communication, 11(10s), 168–176. Retrieved from https://ijritcc.org/index.php/ijritcc/article/view/7616

[16] S. P. Swain, P. K. Swain, "Performance Analysis of a Priority Based Buffer-Aided Relaying," *International Journal of Computer Sciences and Engineering*, Vol.7, Issue.6, pp.1006-1011, 2019.

[17] L. Parida, B. R. Parida, M. R. Mishra, S. K. Jayasingh, T. Samal and S. Ray, "A Novel Approach for Human Activity Recognition Using Vision Based Method," 2023 1st International Conference on Circuits, Power and Intelligent Systems (CCPIS), Bhubaneswar, India, 2023, pp. 1-5, doi: 10.1109/CCPIS59145.2023.10292055.